

CLAIMS

What is claimed is:

1. A method comprising:

receiving a query containing at least one query term;

5 performing at least one of:

(A) determining whether the query includes one or more compound query terms, and if so, automatically expanding the query to include one or more alternative representations of said one or more compound query terms;

10 (B) determining whether one or more query terms are included in a set of inflectional forms, and if so, automatically expanding the query to include one or more corresponding inflectional forms from the set of inflectional forms; and

(C) determining whether one or more query terms are included in a set 15 of alternative spellings, and if so, automatically expanding the query to include one or more corresponding alternative spellings from the set of alternative spellings;

searching a database using the expanded query; and

returning results to a user.

2. The method of claim 1, in which the method includes determining whether the query includes one or more compound query terms, and if so, automatically expanding the query to include one or more alternative representations of said one or more compound query terms.

5 3. The method of claim 1, in which the method includes determining whether one or more query terms are included in a set of inflectional forms, and if so, automatically expanding the query to include one or more corresponding inflectional forms from the set of inflectional forms.

10 4. The method of claim 1, in which the method includes determining whether one or more query terms are included in a set of alternative spellings, and if so, automatically expanding the query to include one or more corresponding alternative spellings from the set of alternative spellings.

15 5. The method of claim 4, in which the method further includes performing (B) and in which automatically expanding the query to include one or more corresponding alternative spellings from the set of alternative spellings is performed before automatically expanding the query to include one or more corresponding inflectional forms from the set of inflectional forms.

6. The method of claim 1, in which the method includes performing at least two of said (A), (B), and (C).

7. The method of claim 1, in which determining whether the query includes one or more compound query terms includes comparing a query term to a list of compound terms.

8. The method of claim 7, in which said one or more alternative representations of said one or more compound query terms are obtained from the list of compound terms.

9. The method of claim 1, in which the query is written in German.

10. The method of claim 1, in which the actions are performed in the order recited.

11. A method comprising:
identifying a set of terms associated with a document;
expanding the set of terms associated with the document by further associating with the document one or more of the following:
one or more alternative spellings of at least one term in the set of terms associated with the document;
one or more alternative representations of at least one compound term in the set of terms associated with the document; and
one or more additional inflectional forms of at least one term in the set of terms associated with the document;
indexing the document using the expanded set of terms.

12. The method of claim 11, further comprising:
receiving a query from a user, the query containing one or more of the alternative spellings, alternative representations, or additional inflectional forms; and identifying the document to the user as being responsive to the query.

5

13. The method of claim 11, in which the document comprises a web page.

14. A method comprising:
searching a first set of documents for hyphenated words;
searching the first set of documents for non-hyphenated words that
10 correspond to said hyphenated words; and
generating a set of associations between said hyphenated words and said corresponding non-hyphenated words.

15. The method of claim 14, further comprising:
searching the first set of documents for pairs of separate words that
15 correspond to the non-hyphenated words and corresponding hyphenated words;
further associating the pairs of separate words with the set of associations between said hyphenated words and said corresponding non-hyphenated words.

16. The method of claim 14, further comprising:
receiving a query from a user, the query containing a first query term;
20 locating the first query term in the set of associations between hyphenated words and corresponding non-hyphenated words; and

expanding the query to include a second query term, the second query term being associated with the first query term in the set of associations between hyphenated words and corresponding non-hyphenated words.

18. The method of claim 14, further comprising:
 - locating a hyphenated word in a document;
 - searching for the hyphenated word in the set of associations between hyphenated words and corresponding non-hyphenated words;
 - if the hyphenated word is not found in the set of associations between hyphenated words and corresponding non-hyphenated words, de-hyphenating the hyphenated word; and
 - indexing the document using the de-hyphenated word.

15 19. A computer program package embodied on a computer readable medium,
the computer program package including instructions that, when executed by a processor,
cause the processor to perform an action selected from the group consisting of:
 expanding a query received from a user by including one or more alternative
 spellings of at least one query term;
20 expanding the query with one or more alternative representations of at least
 one compound query term; and

expanding the query with one or more inflectional forms of at least one query term.

20. The computer program package of claim 19, further including instructions that, when executed by a processor, cause the processor to perform actions comprising:

5 searching a database of documents using the expanded query;
 identifying one or more documents responsive to the expanded query; and
 preparing a list of said one or more documents for transmission to the user.

21. The computer program package of claim 19, further including instructions that, when executed by a processor, cause the processor to perform actions comprising:

10 sending the expanded query to another computer system; and
 receiving from the other computer system a list of one or more documents responsive to the expanded query.

22. An information retrieval system, the system comprising:

15 a document database, the document database containing a group of documents; and
 query processing logic operable to receive a query, expand the query using one or more linguistic techniques, and search documents in the document database for information responsive to the query.

23. The system of claim 22, in which the one or more linguistic techniques comprise one or more of compound term expansion, inflection set expansion, or orthographic expansion.